Environmental Audio Classification Based on Active Learning with SVM*

Yan Zhang
School of Computer and information
Southwest Forestry University
Kunming, China
zydyr@163.com

Danjv Lv
School of Computer and information
Southwest Forestry University
Kunming, China
lvdanjv@hotmail.com

Ying Lin
School of Software
Yunnan University
Kunming, China
linying@ynu.edu.cn

***Abstract*— In order to solve the classification of environmental audio data under the fewer number of the training examples, this paper combined Support Vector Machines SVM) and Entropy Priority Sampling (EPS), and proposed the SVM_EPS method as the selecting sampling strategies in active learning. And the method MOA (Multi-variant Optimization Algorithm) was exploited to select the optimal kernel parameters of SVM. In experiments, the CELP features in 11 dimensions were extracted from the given environmental audio data, and the classification performances were compared under different percent training samples with SVM, EPS and SVM_EPS. The results show that SVM_EPS method outperforms the SVM and EPS.**

Keywords—active learning; environmental audio classification; support vector machines; SVM_EPS; MOA

I. INTRODUCTION

Audio classification is a basis for further audio retrieval and analysis [1]. The environmental audio classification is attracting the attention of researchers increasingly [2-4].The existing techniques for audio classification including minimum distance classifier, neural network, support vector machines [5], decision tree, and hidden Markov Model [6]. The literature [7] realized automatic classification of birds with statistical manifold method to promote the study of ecological environment. Li Yong et.al [8] combined the stream learning with SVM, which obtained the performance of classification efficiently and accurately for ecological environmental audio data. It is difficult to find the optimal classifier with good generalization and to improve the performance of single classifier.

Various classification models have different performances. Training classifier is the key issue in classification research. In the traditional supervised learning, the large number of labeled training examples is required in building classifier model. With the number of the labeled examples decreases in the supervised classification, the performance will get worse. In the actual application, with developing technology of the data collection and storage, it is easy to obtain many unlabeled samples for environmental audio data, while it is too expensive or tedious to acquire the labeled ones. So more research focus on that higher accurate rate can be obtained based on the few labeled and lots of unlabeled examples [9-10]. In the machine learning fields, ensemble learning, semi-supervised learning [11] and active learning [12-13] can effectively deal with this issue. How to use a small labeled data to improve the learning performance becomes in the key problem, which the pattern recognition and machine learning researchers are focusing on.

This work was supported by the national natural science foundation of China under the Grant No. 61462078.

For the environmental audio data with few labeled samples, it is a good way to combine various algorithms and exploit complementary between different classifiers to boost the classification accuracy of environmental audio. The literature [3] exploited ensemble technologies including Bagging, AdaBoost, Random Forests and MCS (Multiple Classifier System), combination of different single classifier. That can obtain better performance than any other single classifier. Semi-supervised learning and active learning, as methodologies of machine learning, make the best use of the unlabeled samples to assist the few labeled examples in establishing classifier model to improve the performance of classification even under the fewer number of the training examples. Zhang Yan et.al proposed employing EPS (Entropy Priority Sampling) and SDS (Simple Disagreement Sampling) methods as the selecting sampling strategies in active learning [4]. For the given environmental audio data, the CELP features in 11 dimensions are extracted. The experiments with the single classifier, EPS and SDS on the environmental audio are carried out in order to illustrate the results of the proposed methods and compare their performance under different percent training sample. The experimental results show that active learning can effectively improve the performance of environmental audio data classification, even under the fewer number of the training examples. The EPS method outperforms the SDS.

Support vector machine (SVM) [14] requires no assumption about data distribution and uses very efficient principles not to over-fitting the test or new data samples. SVM finds an optimal classification hyper-plane through minimizing the upper bound of the classification error. Basically, SVM iteratively located hyper-planes amongst the training data and thereafter optimizes them according to error associates with each variable. It applied in natural environmental sound classification to obtain good generalization [5].

Base on the active learning, this paper mainly focuses on combining multi-SVM classifiers and Entropy Priority Sampling methods to select the most informative unlabeled samples. The behavior and performance of active learning with SVM for environmental audio data classification are explored.

II. ACTIVE LEARNING

A. Active Learning by EPS

In active learning, the most informative unlabeled examples are selected, and used to update the training set on the basis of a supervisor who attributes the labels to the selected samples.

The query function selects samples from the unlabeled pool, which have maximum ambiguity to belong to each class. The supplement labeled training data could boost the performance of the prediction of unlabeled samples to some extent. Fig.1 describes the basic chart flow of active learning [4]. The classifier model can be built based on the training data from the labeled samples. At each iteration, the classifier actively chose the most informative unlabeled examples through the sampling strategy, and submitted them to expert (oracle) to label. The labeled examples are added in training set for next iteration.

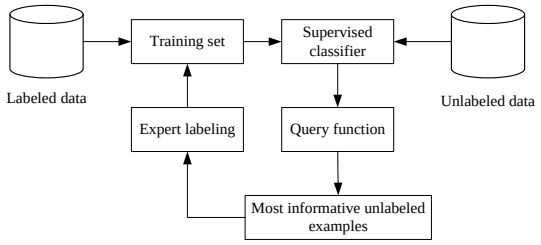

Fig. 1. Description of general active learning

The key point in active learning is sampling strategy [15]. The query by committee is the most simply and common sampling way [16]. Firstly, based on training labeled samples two or more base classifiers are built to form committee. This committee predicted the unlabeled example, then, the most disagreement votes examples are chosen as candidates. Therefore, the most useful for classification examples are supplemented into training set, providing the rich information to improve the efficiency of learning.

Given *k* classifies, the committee-based sample selection techniques are exploited *t* select examples for training. Each unlabeled sample is infused into *k* classifiers, respectively. The point that its prediction results are the most disagree served as the most informative one. Disagreement among the *k* committee members can be measured by the entropy of the classifications voted by each member. For any unlabeled sample, we know that the highest of entropy, the most disagree, the richest information, the largest contribution to classification.

$$Entropy(x) = -\sum_{i=1}^{c} p_i \log p_i \qquad (1)$$

$$p_i = \frac{V(i)}{k} \qquad (2)$$

V(i) is the number of votes of class *i*, *c* is the total number of classes. Examples corresponding to higher entropy have priority of selection over others. The sampling process was named as Entropy Priority Sampling (EPS) [4].

B. Active Learning based on SVM(SVM_EPS)

The support vector machines performs well in common supervised classification method under few EPS method. The SVM serves as base classifier in training data. In order to enlarge difference among classifiers, the bootstrap was taken as resampling method, which resamples the initial training data randomly to consist of various training subset. And different SVM models were built. The detail description of SVM_EPS is shown in TABLE I.

TABLE I DESCRIPTION OF ENTROPY PRIORITY SAMPLING BASED ON SVM

Algorithm: SVM_EPS(SVM_Entropy Priority Sampling)
Input: L: original labeled data, U: unlabeled data,
SVM: learning algorithm,
k: the number of SVM classifier, N:the number of iteration
Output: H_{out}: the final SVM_EPS classifier

Repeat N times
L_a←Ø;
//Resampling, training k classifier based SVM
For t=1 to k
L_t=bootstrap(L); h_t=Train(L_t,SVM);
For each $x_i \in U$
$h_m(x_i)$ (m = 1, 2, . . ., k);
Compute *Entropy*(x_i);
End For
L_a←{x| top n of sort Entropy(x) in desc order, $x \in U$ }
U←U-L_a; L_a←Label(L_a);
$L \leftarrow L \cup L\_a$;
End Repeat
H_{out}←Train(L,SVM)

C. Optimal Kernel Parameters for SVM

Support vector machines SVM [14], established for small training data classification model, is a good method. The main key factor, affecting the performance of SVM, is the appropriate kernel function and the parameters adopted. In generally, four kernel functions are involved in SVM, such as linear kernel function, polynomial kernel function, radial basis kernel function and sigmoid kernel function. In the paper, radial basis kernel function is used. And two optimal parameters including penalty parameter *C* and kernel parameter γ are required to select to boost the performance.

In general, the grid search is adopted to find the optimal *C* and γ . But it will take much amount time to search. In practical application, some heuristic algorithms including generic algorithm and particle swarm algorithm are exploited for SVM parameters optimization. Those methods could find the global optimal solution and they do not need to traverse all possible parameters within the grid. In this paper, a new heuristic algorithm named MOA (Multi-variant Optimization Algorithm) is exploited to find the optimal parameters for SVM kernel function.

MOA [17] is original from the ordered bi-list of computer data structure. In the process of this method, at first, the globe search atom searched in the globe range, then the more detailed search are carried out by local search atom, which could improve the optimization results. Finally, a better solution is recorded in the structural table. Through the global-local

iteration based on structural table, the solution of MOA is recorded in the queue of the structural table. The structural table is listed in Fig.2.

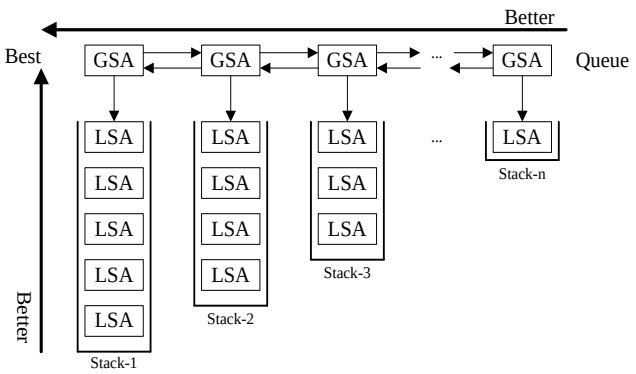

Fig.2.The structural table of MOA

TABLE II DESCRIPTION OF SVM_MOA

Algorithm： SVM_MOA(MOA for SVM)
Input: Q_{depth}:the length of global search atom queue
S_{lenqth} :the length of search atom stack
N_{global}: global search atom
N_{loal}: local search atom
R: search radius
Output: optimal parameters C, γ

Begin
Set the solution space to search C and γ;
Define the fitness function;
Activate the structural table with initial search atom；
Calculate the value of fitness function with SVM classifier model;
While *(Fintness < Max error) && (k<Max iteration)*
Generate randomly the global search atom N_{global};
For each global atom(i=1: Q_{depth})
Inset $N_{global}[i]$ into queue；
Sorted by value of fitness function;
End for
For each local atom(j=1: S_{lenqth})
Insert $N_{local}[j]$ into stack;
Sorted by value of fitness function;
End For
Adjust the global search atom and top stack element of corresponding local search atom;
Sort the queue node on value of fitness function;
End While
Output the optimal solution of C and γ.
End

The following rules should be abided with in the structural table.

1) The global search atom is recorded in the queue node;

2) The queue node is ranked by the value of the fitness;

3) A stack is linked under each queue node, and its depth descended from left to right;

4) The local search atom is generated around the center that is the corresponding solution of queue node, and its order is ranked on the fitness value of the local search atom and corresponding queue nodes in the stack.

In MOA, each search atom represents a potential optimal solution. Search atom is divided into two kinds, global search atom and local search atom. Global search atom is generated with equal probability randomly in the global range, while the local search atom is randomly generated in a certain radius about the queue node of the stack. The search radius is set by MOA parameter. The value of fitness function is up to accuracy of classification, the parameter C and γ of SVM training classification model under 10-folds cross validation are adopted as threshold value. The search process is illustrated in TABLE II.

III. EXPERIMENTAL ENVIRONMENTAL AUDIO DATA

The experimental data are acquired from network and field recording, with 8k sampling rate, 16 bits and mono-track. The environmental audio data includes five classes, such as the sound of different kinds of birds, frogs, wind, rain and thunder. The sound length amounts to almost 10 minutes. Silence and noise are removed in the pre-processing of environmental audio data.

A. The feature extraction of environmental audio

The feature extraction is executed based on the bit-stream through the G.723.1 data encoding on the Matlab platform. CLEP (Code excited Linear Predictive) is characterization by coefficient of short tube cascade channel model. CELP features are mainly from LPC, Linear Prediction Coefficient analyses the sound mechanism from the original source. Through the short tube of channel cascade model research, the system transfer function is in line with the pile in the form of digital filter, so the signal of t time can be used several times before combination of the signal to estimate. By making the actual speech samples values and linear prediction to achieve minimum mean square error between the sampling value LMS. The linear prediction coefficients can be obtained, that is LPC.

LPC and pitch features are extracted at each bit-frame after the unpacked bit-stream. 10 order coefficients of LPC is obtained at each bit-frame, from 0 ~ 23 bit (LPC0~LPC2), which consist of the 10 dimensions of LPC features. 1 dimension pitch is extracted from the 24th to 38th bit at each bit-frame. Finally, the 11 dimension features of CLEP (Code excited Linear Predictive) are composed. The features extracting process is shown in Fig.3.

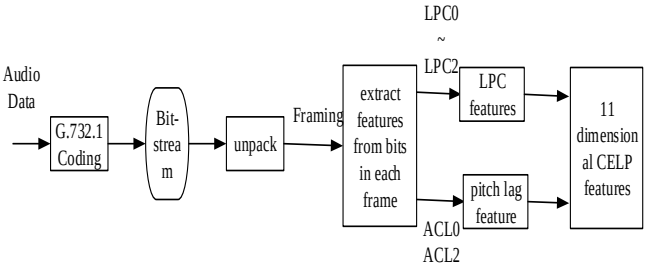

Fig.3. Extraction of CELP environmental audio features

B. Experimental Data

The frame of the experimental audio is he basis unit for data statistics and classification in the experiments. In order to

avoid large amount of data overflowing in training, the experimental data is selected from the total number of frames in the sampling according to one-third of the total data in each category. According to the proportion of 3:1, the samples are divided into training data set and test set, that is, 75% as training samples and 25% as testing samples. The samples in two sets keep the same distribution. Five classes of audio signal frame and class labels of samples are shown in TABLE III.

TABLE III THE DATASET IN EXPERIMENT

class	Frames	Frame/time	75%-training	25%-test
bird	3900	3000	2250	750
wind	6434	2000	1500	500
rain	6200	2000	1500	500
frog	3334	1200	900	300
thunder	1634	1000	750	250
total			6900	2300

In experiment, the rate of training examples extracting from labeled samples are 10% ,20%,40%,60% and 80%,respectively. In order to verify the validity of the method, it is 10 times sampling that each time training data are in proportion to be taken, randomly. The experimental results are taken the average of 10 group data.

IV. RESULTS AND DISCUSSION

In the experiments, the performance of traditional supervised single classifier and active learning classification model are compared. Four different classification algorithms are used in single classifier, including decision tree (J48), Naïve Bayes (NB), Radial Basis Function (RBF) and Support Vector Machines (SVM).For SVM, when the classification model was established based on training data, SVM_MOA method is exploited to choose the optimal parameters C and γ of kernel function. In SVM_MOA, the queue length and the number of global search atom are set 6, the search radius is 0.01, and the dimension of parameters is 2. Active learning EPS using J48,NB and RBF to choose the most informative sample points, while active learning SVM_EPS with six SVM classifiers to calculate the vote entropy for each unlabeled example. Six SVM classifiers involved in SVM_EPS are different classifier model, whose initial training data are random sampling through bootstrap method. Under different rate of training samples, different classification error rate of various methods are listed in TABLE IV.

TABLE IV ERROR RATE OF DIFFERENT CLASSIFICATION METHODS

Training Samples	Single classifier				Active learning	
	J48	NB	RBF	SVM	EPS	EPS_SVM
10%	0.2159	0.2271	0.1823	0.1652	0.1709	0.1509
20%	0.1984	0.2216	0.1812	0.1526	0.1594	0.1478
40%	0.1810	0.2111	0.1866	0.1287	0.1495	0.1230
60%	0.1653	0.2062	0.1851	0.0996	0.1474	0.0987
80%	0.1659	0.2086	0.1875	0.1087	0.1417	0.1039

According to results from TABLE IV, for a single classifier, the performance of SVM is superior to the other traditional supervised classification methods. For example, under different training samples rates, related to the J48 classifier, the performance of SVM increased by 23.48%, 23.08%, 28.90%, 39.75% and 34.48%. And to the EPS method with three different classifiers, that of SVM increased by 3.34%, 4.27%, 13.91%, 32.43% and 23.29%, respectively. The SVM_EPS, applying the SVM classifiers into EPS, its error rate of classification is lowest. To the EPS, the performance increased by 11.70%, 7.28%, 17.73%, 33.04% and 26.68%, while related to SVM, that increased 8.66%, 3.15%, 4.43%, 0.90 and 4.42%, respectively. In view of traditional single supervised classifier, active learning method exploits the characteristics of unlabeled samples, during the training iteration process and enlarges the differences among base classifiers, which improved auxiliary generalization and accuracy of classification under small training data.

Fig. 4 shows the classification error rate comparison among SVM, EPS and SVM_EPS. SVM_EPS combines the SVM and EPS, adopts multiple SCM classifiers and ensembles advantages of SVM and active learning EPS. Therefore, SVM_EPS outperforms the other. Especially, under the few training samples, such as 10% training data, the performance of SVM_EPS improved significantly compared with SVM. In the case of 60% to 80% training samples, with the increase of the number of samples, the performance of SVM_EPS decreased. There maybe have two reasons. One is that training data only removed the simple mute and noise, and do not purify the data for training the classification models. The other is in the process of active learning iteration, the unlabeled examples will be mistakenly labeling, and the noise data are put into the training set to increase the error of samples.

In addition, the heuristic multi-variant optimization algorithm MOA is exploited to fine the optimal kernel function parameter of SVM. After each iteration, the training samples are added, and the retraining the SVM classifier should need to choose new kernel parameter. So the efficiency of SVM_EPS is lower than that of EPS.

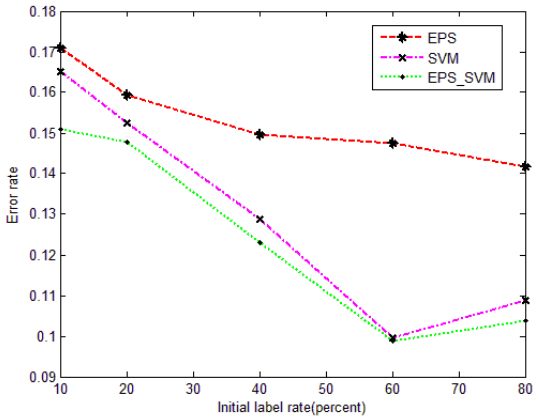

Fig.4. Classification performance of three methods

V. CONCLUSION

For the environmental audio data, the cost of labeling the samples is higher and tedious. In the case of small training data, support vector machines (SVM) is a better classification method. Active learning exploits a lot of unlabeled examples, selects the most informative sample points with significant disagreement to increase the difference of base classifiers, and enlarges the number of training data, which can improve the performance of classification. This paper combined the SVM classifier and active learning EPS method, and different SVM classifiers with different training set are involved in EPS. The selected unlabeled examples with higher vote entropy are a beneficial supplement to the training set. According to the analysis of experimental result data, SVM_EPS outperformed SVM and EPS in classification performance and generalization.

The features of data samples are also one of the main factors to affect the classification performance. In this paper, the CELP audio characteristics of environmental audio are extracted in classification process. The next further research work will extract more effective multi-view features from the structure of audio data, combine feature space and classification algorithms. The more effective and efficient classification strategy will be focus on to improve the environmental audio classification accuracy.

REFERENCES

[1] Bai Liang, Lao Song-Yang, Chen Jian-Yun et.al., Audio Classification and Segmentation Based on Support Vector Machines [J]. Computer Science.2005, 32(4):87-90

[2] Xiao-mei Zhang, Ding-cai Yang. Environmental audio classification based on support vector machines. Electronic Measurement Technology. 2008,31(9):121-123.

[3] ZHANG Yan, LV dan-jv, WANG hong-song. The application of multiple classifier system for environmental audio classification. ICMIT 2013: proceedings of the 2013 International Conference on Mechatronics and Information Technology, Guilin, October 19-20, 2013[C]. Applied Mechanics and Materials, 2013.

[4] Zhang Yan, Lv Dan-Jv,Wang Hong-song. Research of Environmental Audio classification Based on Active Learning [J].Computer technology and development, 2014,24(6):110-113.

[5] Yu Qing-Qing, Li Ying, Li Yong. A SVM-based classification Approach for Natural Sounds [J]. Computer& D.2010,38(7):1-5.

[6] Yi-wen Zheng. Typical Methods for Audio Classification. Ji Suan Ji Yu Xian Dai Hua. 2007,(8):59-63.

[7] Briggs F, Raich R,Fern X Z. Audio Classification of Bird Species:A statistical Manifold Approach[C]//Proc. Of the 9th IEEE International Conference on Data Mining, 2009:51-60.

[8] Li Yong, Li Ying, Yu Qing-Qing. Environmental Sound Classification Based on Manifold Learning and SVM [J]. Computer Engineering.2011,37(7):288-290.

[9] Zhou Zhi-Hua, Wang Yu. Machine Learning and Application [M]. Beijing: Tsinghua University Press,2007.

[10] Seeger,M. Learning with labeled and unlabeled data, Technical Report[R]. University of Edinburgh, Edinburgh, UK, 2001.

[11] Chapelle O, Scheolkopf B, Zien A. Semi-Supervised Learning [M]. Cambridge: MIT Press, 2006.

[12] Long Jun, Yin Jian-Ping, Zhu EN. A Survey of Active Learning [J]. Journal of Computer Research and Deveopment.2008,45(suppl.):300-304.

[13] Liu Kang, Qian Xu,Wang Zi-Qiang. Survey on Active Learning Algorithms [J]. Computer Engineering and Applications. 2012, 48(34):1-4.

[14] Vapnik V. Statistical Learning Theory [M].Wiley, New York.1998.

[15] Wu Wei-ning, Liu Yang,Guo Mao-Zu,et al. Advances in Active Learning Algorithms Based on Sampling Strategy [J]. Journal of Computer Research and Development. 2012,49(6):1162-1173.

[16] Seuong,H., Opper,M., Sompolinski, H. Query by committee[C]. In: Proceedings of the 5th ACM workshop on computational learning theory, Pittsburgh, PA, 1992, 287-294.

[17] Li B L,Shi X L, Gou C X, et al., Multivariant Optimization Algorithm for Multimodal Optimization[C].Applied Mechanics and Materials, 2014,483: 453-457.